DOCUMENT RESUME

ED 192 456 EA 012 949

AUTHOR Airasian, Peter W.: And Others
TITLE Concepts of School Effectiveness as Derived from Research Strategies: Differences in the Findings.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Nov 79
GRANT NE-6-00-3-0214
NOTE 50p.: Tables may not reproduce clearly due to small print of original document.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement: Concept Formation: Data Analysis: Data Collection: *Educational Assessment: Elementary Secondary Education: Equal Education: *Evaluation Methods: Family Environment: Models: Predictor Variables: *Research Methodology: *Research Problems: Resource Allocation: Statistical Analysis: Student Characteristics: *Student School Relationship
IDENTIFIERS *Coleman Report: *School Effectiveness

ABSTRACT

The methodological parameters of evaluative research studies help to identify a conceptualization of the phenomenon under investigation. Holistic critiques that identify the conceptual models under which the research was conducted shed light on the validity of the disparate inferences made from a number of studies of schools and school effectiveness. This paper attempts such a critique of studies of school effectiveness in general, and the 1966 "Equality of Educational Opportunity Survey" by Coleman and others in particular. A table presents a brief overview of three studies of school effectiveness. Differences in the methodological parameters (data analysis strategy, stratification levels, level of data aggregation, and selection of independent and dependent variables) and the impact of these differences on conclusions about school effectiveness are identified and discussed. It is important that discussion of the methodological parameters characterizing varied research studies be put in a more general, conceptual framework, so that the model of schooling being examined by different researchers can be identified and critiqued. (Author/MLF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192456

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Concepts of School Effectiveness

Peter W. Airasian
Thomas Kellaghan
George F. Madaus

EA 012 949

The National Institute of Education
U.S. Department of Health, Education and Welfare
Washington, D.C. 20208

2

This report was prepared under NIE Grant # NE-6-00-3-0214 from the National Institute of Education, Testing, Assessment and Evaluation Division, U. S. Department of Health, Education and Welfare. The opinions expressed in this report do not necessarily reflect the position or policy of the National Institute of Education or the Department of Health, Education and Welfare.

This report was drawn from a larger study supported by the Carnegie Corporation of New York, the National Institute of Education, the Russell Sage Foundation, and the Spencer Foundation.

November, 1979

Concepts of School Effectiveness as Derived from Research Strategies: Differences in the Findings

Peter W. Airasian, Boston College
Thomas Kellaghan, Educational Research Centre,
St. Patrick's College, Dublin
George F. Madaus, Boston College

Evaluative researchers face a number of methodological decisions in carrying out their studies. They must determine the level to which data will be aggregated, the nature of the independent and dependent measures, the appropriate stratification levels and the strategy of data analysis. Even in a similar problem area different evaluators make quite different decisions regarding such methodological parameters. For example, a review of studies of school effectiveness shows that data have been variously aggregated to the school level (Mayeske, *et al,* 1972), the individual pupil and school level (Coleman, *et al,* 1966) and the class within school level (Madaus, Kellaghan and Rakow, 1975). Analyses have been stratified by social class (Frederiksen, 1975), geographic region and racial-ethnic group membership (Coleman, *et al,* 1966) and not stratified at all (Mayeske, *et al,* 1972). Similarly, different studies have used different methods of data analysis and different types of independent and dependent variables.

In considering methodological differences among studies which purport to examine school effectiveness, we must recognize that there are, of course, no necessarily right or wrong choices among alternative methodological options. Nonetheless, the methodological choices an evaluator makes are important because, taken together, they help to define the *de facto* conceptual framework within which to examine the results of his particular study and to make comparisons of that study with other studies. Recognition of the *de facto* conceptualization in evaluative research studies can help make crucial distinctions among and syntheses of different studies in a given problem area.

If maximum benefit is to be derived from evaluative research studies, it is necessary to examine and critique them holistically. Here 'critique' means more than to criticize.

Critique is used in the classical sense of exposing the essence or identifying the inherent meaning. While the specific methodological shortcomings of evaluative research studies need criticism, a more holistic critique based upon the *de facto* conceptualization of school effectiveness exemplified by the research should not be slighted in a spate of specific methodological criticism. Holistic critiques which identify the conceptual models under which the research was conducted shed light on the validity of the disparate inferences made from a number of studies of schooling and school effectiveness.

This paper attempts such a critique of studies of school effectiveness in general, and the *Equality of Educational Opportunity Survey* (Coleman et al, 1966) hereafter called the Coleman Report, in particular. Our intent is to show how different conceptualizations of school emerge from the use of different methodological parameters such as the level of data aggregation, measures of independent variables, and models of data analysis and to show further how altering one or more of these parameters can alter dramatically inferences made about schooling and school effectiveness.

The Coleman Report

By nature of its Congressional mandate, the central focus of the Coleman Report was an examination of differences in school resources and achievement among racial and ethnic groups in American public schools. The Report addressed four key issues. First, it sought to determine the extent of racial and ethnic segregation in the schools. Second, it surveyed by geographic region, the facilities and resources available in schools attended by different racial and ethnic groups; it sought to determine the degree to which the nation's schools provided racial and ethnic groups equal educational opportunities in terms of facilities and resources. Third, it measured student achievement using standardized tests. Finally, it sought to determine the relationship between school resources and facilities and student achievement.

To address these issues, the largest survey of American public education in history was undertaken. Nationwide, 645,000 pupils in grades one, three, six, nine and twelve were administered standardized tests of ability and achievement. Sixty thousand teachers completed questionnaires about their educational backgrounds, experiences and attitudes. Data were collected from over 4,000 schools on factors ranging from the average teacher salary to the age of textbooks in particular subject areas.

The two volume report produced the following four major findings:

1. A decade after the Brown decision which outlawed separate but equal education, racial separation was still widespread in American schools.
2. Within eight geographic regions of the country, the physical facilities, curricula and teacher characteristics were quite similar in schools classified as either predominantly white or predominantly black.
3. Standardized achievement and ability test scores were higher for white pupils than for most racial and ethnic minority groups.
4. There was very little relationship between differences in the physical facilities and resources present in schools attended by diffferent racial and ethnic groups and the achievement of pupils in those schools. Analyses showed that the largest achievement differences occurred among pupils in the same schools rather than among groups of pupils attending different schools with different resources and facilities. Further, the small differences that were found in the achievement of pupils in different schools were largely related to home background or social class factors rather than to any differences in school resources or facilities.

The Coleman Report precipitated a heated and prolonged debate about the differential effectiveness of schooling. A study which had set out to document for Congress inequalities in school resources and facilities, so that remedial

legislation could be drafted, ended up by concluding not only that resource and facility differences between black and white schools within geographic regions were not large, but further that resources and facilities didn't appear to make much difference to student achievement once home background factors were taken into account. Only about ten percent of the total variation in pupils' achievement was found to lie between schools. Of this ten percent, only a small amount, on the order of one to two percent, was related to differences in school resources and facilities.[1] The finding that school characteristics accounted for very little of the large differences in pupils' achievement was taken to indicate that practice in a major social institution was not nearly as effective as generally believed. The popular press was more direct, suggesting that 'schools don't make a difference.'

The conclusions that have been drawn from the Coleman Report are often diametrically opposed. At one extreme, it is concluded that schools are not important since they contribute nothing or very little to student achievement. At the other extreme, the research findings are regarded as defective and counter-intuitive, contradictory to a basic belief that schools and their resources play an important role in the development of individuals and in the attainment of more general social policies such as the reduction of poverty.

Undoubtedly the Coleman Report was flawed in a number of respects, many unavoidable, given the combination of a lack of previous experience in conducting and managing large-scale, policy-oriented research, the short lead time for planning, and the mailed survey technique used to gather much of the data. The data base has been criticized on a number of counts: the low and selective return rate for questionnaires (Armor, 1972; Bowles and Levin, 1968; Shea, 1976), the lack of sample representativeness (Hanushek and Kain, 1972), and the failure to validate questionnaire data

[1]These findings were true for white, Northern pupils. Generally higher percentages of between school variance were found for other racial groups in other geographic regions. School-related factors also tended to account for slightly higher proportions of the observed between school variance for these groups.

with observation or follow-up studies (Jencks, 1972; Mosteller and Moynihan, 1972). Other critics have pointed out that the Coleman Report was a survey, not an experiment, and therefore was able to produce only relational inferences about school variables and pupils' achievement. The Coleman Report provides no evidence about what would happen to pupils' achievement if the school resources surveyed were actually manipulated (Mosteller and Moynihan, 1972; Mayeske, *et al,* 1972). These criticisms are valid. However they represent problems which are endemic to large scale evaluative research and do not speak to the conceptualization of schooling implicit in the Report. It is the Report's conceptualization of schooling, generally unarticulated and frequently unrecognized, which is at the heart of the subsequent debate and controversy about school effectiveness.

It must be emphasized that when we talk of the conceptual model of schooling which underlies the Coleman Report, or other research, we are not referring necessarily to an explicit statement made by the researcher. Rarely do evaluative researchers make explicit the premises of conceptualizations which guide their studies. Often there are no prevailing, widely accepted conceptualizations to guide study or, conversely, there is a plethora of conceptualizations of equal credibility. In light of this fact, the reader must recognize that the conceptualizations of which we speak are *de facto* conceptualizations, which can be identified from a consideration of the research methodology. In particular, the type of independent and dependent variables measured, the level to which data are aggregated for study, the selection of stratification variables, and the data analytic strategy can be used to identify a *de facto* model of schooling. The identification of this model is the first step in any holistic critique of an evaluative research study.

In considering what the methodology of the Coleman Report suggests about a conceptualization of schooling and school effectiveness, two caveats should be kept in mind. First, the Coleman Report was a first generation study. There is little doubt that the researchers strongly felt the lack of pre-existing conceptions of school learning. The research

team had little to guide it in the selection of specific variables and methodology except for a Congressional mandate to focus on differences in equality of educational opportunity offered by schools. Second, the sheer size of the study and the time constraints imposed by Congress resulted in many purely pragmatic decisions. For example, it would have been difficult to construct and validate a new series of achievement measures, so existing standardized tests were used. Further, the school factors measured were selected largely on the basis of their being amenable to manipulation by policy makers (Coleman, 1975).

Three conclusions can be made regarding the conceptualization of schooling embodied in the Coleman Report. First, from the data aggregation procedures, one can conclude that the primary effect of schooling on student achievement was presumed to occur at the level of the school itself; the effect of schooling on subgroups of pupils, teachers, classes, tracks or streams within the school was perceived to be less important than a more general, school-wide effect. Second, the primary measures of school quality were status variables[1] associated with teacher experience, age of textbooks, curricular offerings, presence of laboratories, and so forth. Such independent variables represent static qualities of schools and teachers. Third, the selection of a standardized, norm-referenced ability test as the dependent variable implies that the researchers presumed that such tests were valid and sensitive measures of the cognitive outcomes of schooling. Taken together, these three observations represent the following conceptualization of schooling: The school as a whole, by virtue of its static resources and facilities, influences pupils' general cognitive outcomes as measured by commercially available standardized tests. This relatively simple conceptualization, stripped of statistical analyses, procedures and tables, represents the essential view of schooling implicit in the Coleman Report.

---

[1]Many school effectiveness researchers and their critics commonly refer to the independent variables measuring school resources and 'input' variables. Notice that the initial selection of this term itself denotes a great deal about the conceptualization of schooling implicit in researchers' minds. The use of the words 'input' and 'output', particularly in early studies of school effectiveness, suggests a factory, assembly line view of the educational process (Levin, 1970). With time, these words have become ingrained in the school effectiveness parlance and are now commonly used and accepted.

Examination of similar studies or reviews of studies carried out at about the same time as the Coleman Report, does not reveal a conceptualization of schooling which differs essentially from that implicit in the Coleman Report (e.g., Averch, *et al,* 1972; Benson, *et al,* 1965; Burkhead, Fox and Holland, 1967; Goodman, 1959; Kiesling, 1969; Madaus, Airasian and Kellaghan, in press; Mollenkopf and Melville, 1956; Purves and Levine, 1975). Further, these studies, as well as re-analyses of the Coleman *et al* data (Jencks, *et al,* 1972; Mosteller and Moynihan, 1972; Smith, 1972), tended to corroborate the Report's basic inference that schools bring little to bear on student achievement that is independent of home background. National evaluations of Head Start and Title I programs, although somewhat different in focus, shared many of the same methodological characteristics (Madaus, Airasian and Kellaghan, in press) and produced similar inferences about school effectiveness (Cicirelli, *et al,* 1969; Glass *et al,* 1970).

In what follows we shall discuss the effect that a change in the 1) method of data analysis, 2) level of data aggregation, 3) selection of a dependent variable and 4) selection of the independent variables has on inferences about school effectiveness. Our aim is to illustrate how a change in the methodological parameters of school effectiveness studies alters the conceptualization of the schooling process and, in turn, conclusions about school effectiveness.

While the four methodological parameters are treated separately, in practice they overlap and interact to affect the inferences made from evaluative research studies. It often is difficult to tease out the unique impact that changing a single parameter has on inferences about school effectiveness. This difficulty is due largely to the fact that in most studies, or replications, no systematic attempt was made to alter a single parameter, such as the level of data aggregation, and to observe its effect on the inferences made about school effectiveness, holding all other parameters constant. Instead subsequent studies have altered several parameters so that the unique impact of altering one parameter on inferences about school effectiveness is not clear.

Table 1 compares the methodological parameters, conceptualizations and major conclusions of three studies of school effectiveness. The table provides an overview of major methodological differences among the studies and how differences in methodological parameters relate both to a study's conceptualization of schooling and its conclusions about school effectiveness. The three studies outlined obviously are not exhaustive of the many school effectiveness studies done in the last decade or so. However, the three selected studies do present very different conceptualizations of schooling and school effectiveness. Other studies, not listed in Table 1, will be discussed in succeeding sections. It is important to note that conceptually, and to a lesser extent methodologically, these other studies typically are akin to one of the three "ideal type" studies included in Table 1. Table 1, therefore, provides a framework for the following discussion, which will consider the import and influence of the five methodological parameters on school effectiveness research.

Method of Data Analysis

Of the many criticisms directed at the Coleman Report, limitations in the strategy of data analysis have received the greatest amount of attention. There are two features of Coleman, *et al's* (1966) method of analysis which require comment. First, in fulfilling their Congressional mandate to conduct a survey "concerning the lack of availability of equal educational opportunity by reason of race, color, religion or national origin in public educational institutions. . ." (Civil Rights Act of 1964, Section 402) Coleman, *et al* (1966) chose to stratify their sample by geographic region and racial-ethnic group membership. Twenty stratified groups were established: black students in eight geographic regions, white students in eight geographic regions and four other minority groups examined nationwide. Data analyses were performed *separately* for each stratum; amounts of between and within school variation were determined for each stratum. The relationship between school resources and pupil achievement was also investigated within each stratum. Therefore, school

Table 1

Comparison of the Methodological Parameters, Implicit Conceptualizations and Major Conclusions of Three Studies of School Effectiveness

	The Coleman Report (1966)	Mayeske, *et al* (1972)	Madaus, Kellaghan and Rakow (1975)
Method of Data Analysis	Multiple regression with a prior ordering of variables	Commonality analysis	Commonality analysis
Major Stratifications in Data Analysis	Geographical region and racial-ethnic group	No stratification; whole country.	No stratification; whole country.
Level of Data Aggregation	School and individual	School	Classroom within school.
Independent Variables	Predominantly status measures of pupils, teachers, schools, etc.	Predominantly status measures of pupils, teachers, schools, etc.	Predominantly process or climate measures of schools and classes.
Dependent Variable	Standardized, norm-referenced ability test	First principal component of standardized, norm-referenced ability and achievement tests	Curriculum specific external examinations and standardized, norm-referenced ability and achievement tests.
Conceptualization of Schooling and School Effectiveness	The school as a whole, by virtue of its static resources and facilities, influences general cognitive outcomes of the type measured by standardized tests. Home background influences are prior to and independent of school influences.	The school as a whole, by virtue of its static resources and facilities, influences general cognitive outcomes of the type measured by standardized tests. The school's influence is not necessarily independent of pupils' home background characteristics.	Classes within schools, by virtue of their climate and processes, influence cognitive outcomes of the type measured by instructionally specific examinations. Classroom influences are not necessarily independent of pupils' home background characteristics.
Major Conclusions	1. About 10 percent between school variance observed.	1. About 35 percent between school variance observed	1. About 35 percent between class variance observed on both types of dependent variables.
	2. Between school variance explained primarily by home background factors.	2. Between school variance explained primarily by shared home-school component.	2. Between class variance on external exams explained primarily by class press and climate variables; classroom press and climate explained very little of the between class variance on standardized tests.

effectiveness in the Coleman Report was not examined across all American schools or even across all the pupils in a single school, but rather within various racial-ethnic groups and geographic regions. Between school variance attributable to differences in the performance of schools in different geographic regions or schools comprised of different racial and ethnic groups was lost because such groups were studied separately (Smith, 1972). It is interesting to note that in nearly all cases, the proportion of between school variance observed for the four racial-ethnic groups which were studied nationwide exceeded that of the groups studied within geographic regions. Moreover, test performance of the average minority pupil was below that of the average white pupil, implying that had analyses been performed across racial-ethnic strata, the proportion of between school variance observed would have been larger than that reported for separate strata. However, the ten percent between school variance associated with Northern white pupils became the benchmark finding of the Coleman Report; most discussions cite this figure as indicative of the proportion of between school variance. Thus. in the Coleman Report, the conceptualization of school effectiveness did not extend to all American schools regardless of their location or racial-ethnic composition, but was instead limited to pupil groups of particular racial-ethnic composition within different geographic locations.

A second aspect of Coleman, *et al's* (1966) strategy of data analysis has provoked a great deal of criticism. The Coleman Report used an asymmetric variance decomposition procedure to investigate the relationship of school factors to achievement (Bowles and Levin, 1968; Levin, 1970); Madaus, Kellaghan and Rakow, 1975; Mayeske, *et al,* 1972, 1973; Wiley, 1973). Data were analyzed by means of multiple linear regression, a technique that attempts to analyze the collective and unique contributions of two or more independent variables to the variation in a dependent variable (Cohen and Cohen, 1975). Problems occur when a multiple regression analysis is used with a large number of independent variables which are highly co-linear, that is, correlated among themselves. For example, one would expect that the

social class of students in a school will be related to such school variables as per pupil expenditure, class size and pupils' vocational aspirations (Kellaghan, 1976). In an effort to make data analysis more parsimonious and meaningful, Coleman, *et al,* (1966) collapsed the independent variables into four groups or blocks: (1) home background, (2) student body characteristics, (3) school facilities and curriculum, and (4) teacher characteristics.

In the analysis seeking to determine the contributions of each of these blocks to variation in pupils' measured verbal ability, Coleman, *et al* first entered into the regression equations variables representing the students' home background, assuming that the students' backgrounds were 'prior to and independent of, any influence from school factors' (p.330). This rationale for the order in which independent variables were studied is one of the few places in the Coleman Report where a guiding premise related to the researchers' conceptualization of the schooling process was explicitly stated. The result of this decision was to limit the detectable effects of school variables (e.g., teacher facilities and curriculum) to that variation left after accounting for variation in family background characteristics. In effect, by entering the family block variables first in the analysis, these variables were assigned the explanatory power they possesed uniquely *plus* the explanatory power they shared with other, unremoved blocks of variables. By selecting an ordered, asymmetric variance decomposition procedure, Coleman, *et al,* (1966) tended to accord greater weight to the effect of home background variables on the dependent variable, thereby reducing the likelihood of school effects being demonstrated (Bowles and Levin, 1968; Hanushek and Kain, 1972; Wiley, 1973).

To overcome problems associated with asymmetric variance decomposition, a method of symmetric variance decomposition was developed after the Coleman Report had been completed (Beaton, 1973, 1974; Mayeske, et al, 1972; Smith, 1972; Wisler, 1974). This approach, called commonality analysis, partitions 'explainable variance in a dependent variable into the portions attributable to each independent variable and to all combinations of independent

variables' (Beaton, 1974, p. 63). Commonality analysis examines all possible orders of entering blocks of variables into the regression equation and thus the partitioning of variance is not dependent upon the order in which variables are entered. Symmetric variance decomposition procedures can provide an estimate of the variance explained by the home block and the school block acting independently *and* in concert. That is, variance is partitioned into three, rather than two disjoint components: (1) home background acting alone, (2) school factors acting alone, and (3) the joint effect of home and school factors. Commonality analysis gives an estimate of the complex relationship between home, school and pupil performance. The method does not require assumptions about the temporal effect of home and school variables.

Smith (1972) reanalyzed portions of the Coleman Report data to examine the unique and shared contributions of school specific blocks of independent variables to between school achievement variance. However, Smith did not include home background variables in his study and so did not provide estimates of unique and shared variance among both home and school factors.

Mayeske, *et al* (1972) were the first to reanalyze the Coleman Report data using commonality analysis. As Table 1 shows, Mayeske, *et al* (1972) made three additional changes in the methodology of the Coleman Report which, to some extent, cloud unequivocal interpretation of the singular effect that changing the strategy of data analysis can have on evaluative research findings. First, unlike the Report, in which school and teacher variables were aggregated to the school level but pupil home background and test performance variables were kept at the individual pupil level in the analysis (Coleman, *et al,* 1966, Appendix 9.8, p. 571), Mayeske, *et al* (1972) aggregated all data to the school level for analysis. Second, Mayeske, *et al* (1972) did not stratify schools by either racial-ethnic composition or geographic region. Instead, racial-ethnic composition was made an independent variable in analysis. Third, instead of using a measure of verbal ability as the dependent variable, Mayeske, *et al* (1972) used a composite measure correspond-

ing to the first principal component of a factor analysis of all the standardized tests examined in the Coleman Report.

Although both the Coleman Report and Mayeske, *et al* (1972) analyzed the same data set, Table 1 shows that the conceptualization of school effectiveness and pupil achievement differences had changed in the six years intervening between the two studies. We may hypothesize that the findings of the Coleman Report with respect to the percent of between school variance observed (particularly for Northern whites) and the proportions of this variance which were related to school factors, were so devastating in many quarters that in succeeding research, attention was directed at demonstrating that schools do, in fact, differ in substantial ways. Thus Mayeske, *et al* (1972) indicated that a primary question in their reanalysis of the Coleman Report's data was "How do the schools' characteristics influence such things as the achievement level of *all* the students in schools?" (Mayeske, *et al,* 1972, p. 7). Schools were studied as intact units, regardless of pupils' racial-ethnic membership within schools, and more importantly, schools nationwide, unstratified by geographic region were examined. School effectiveness and the school's influence on pupils thus were conceptualized quite differently by Mayeske, *et al* (1972) than they were in the Coleman Report.

With these changes in the methodological parameters, Mayeske, *et al* (1972) found that nationwide, about 35 percent of the variation in pupil achievement occurred between schools. This figure was about three times as high as that reported for Northern whites in the Coleman Report. When the between school variance was partitioned into home, school, and joint home-school components, the following conclusions were reached:

> . . .the influence of the school cannot be separated from that of the student's social background—and vice versa. Moreover, the common influence of the school and the student's social background exceeds either of their distinguishable influences (Mayeske, *et al,* 1972, p. ix).

Approximately 80 to 85 percent of the between school variance was accounted for jointly by home-school

influences. This conclusion about the effect of home and school factors on achievement was quite different than the Coleman Report's conclusion that schools bring little to bear on a student's achievement, after home background factors are taken into account.

It seems reasonable to suggest that the use of the school as the level of data aggregation, the selection of a different dependent variable, and the application of unstratified analyses are most powerful in explaining the increased percentage of between school variation observed by Mayeske, *et al* (1972), and the different inferences about school effectiveness. The decision to use these methodological parameters suggests a more generalized conception of schooling and school effectiveness than that in the Coleman Report — all schools in the country were included and racial and ethnic strata within schools were considered an independent variable. It also seems reasonable to conjecture that the primary effect of using commonality analysis was upon the conclusions Mayeske, *et al* (1972) reached about the contribution of school, home, and joint home-school factors to pupil achievement. The simple availability of commonality analysis undoubtedly did much to color and shape the researcher's inferences about schooling, insofar as the technique was capable of providing an estimate of the interactive influence of home and school factors on pupil achievement. Certainly, commonality analysis removed the necessity of making a priori decisions about the order in which blocks of independent variables would be examined. School influences did not have to be viewed as independent of home influences by Mayeske, *et al* (1972). Thus, the use of a data analytic model different from that used in the Coleman Report and the different conceptualization of the schooling process permitted by commonality analysis led to quite different inferences about school effectiveness. Schools were seen to be effective, but the influence of school factors was intimately intertwined with the student's social background.

Level of Data Aggregation

Of necessity, aggregated data are dealt with in all studies of school effectiveness (Burstein, 1976) and while there are no necessarily right or wrong levels to which data may be aggregated, the selection of a particular aggregation level can mask the effects of schooling at other aggregation levels. Moreover, the selection of one aggregation level over competing levels has a powerful influence on the conceptualization of school effectiveness adopted by a researcher. A crucial question then becomes, how does the selection of one level of data aggregation over other, potentially viable levels influence conceptualization of and conclusions about school effectiveness?

As Table 1 indicates, the Coleman Report utilized two levels of data aggregation. First, teacher, school and some student independent variables were aggregated to the school mean score level. Second, individual pupils' standardized test scores were retained as the dependent variable measure; test performance was not aggregated across individual pupils. There were two reasons for this strategy. First, given the racial-ethnic strata within which analyses were performed, it was necessary to separate the test performance of students of differrent racial and ethnic groups in the same school. More importantly, the Coleman Report was specifically directed towards examining and estimating various components associated with within school variance.[1] If pupil test data were aggregated to the school or even racial-ethnic group within school level, within group variance estimates would have been lost.

Using this strategy of data aggregation the Report indicates that the largest proportion of variance in pupils' test performance lay within, rather than between, schools. Further, very little of the small amount of between school variance that was identified could be attributed to resource

[1]The strategy of analysis used by Coleman, *et al* (1966) and the procedures used to estimate between school and within school variance components is not well detailed in the Coleman Report itself. However, Smith (1972) presents a discussion of the variance partitioning techniques used by Coleman, *et al* (1966). In particular, see the footnotes to Table 6 in Smith.

and facility differences among schools. However, given the way the data were aggregated these results should not have been surprising.

Pupils come to school with different degrees of learning readiness. Once in school, pupils tend to be treated differently according to their readiness level. This differential treatment in turn accentuates the individual differences among pupils within a school. Schools accept a principle of equality *concern* for all pupils in which equal values but not identical components of school resources are allocated to each child. The differential treatment begun in the primary grades generally continues through elementary and secondary grades (Brimer, et al, 1977). If this portrait of how schools operate is accurate, then the Coleman Report's finding that most variance lies within rather than between schools should not come as a surprise. Further, if this portrait is accurate, it also suggests that the general, within school variance component could be broken down further into a component that lies between classes, instructional groups or tracks within schools, as well as a "true" component that lies among individual pupils. The achievement of pupils may vary because of differences in some school-wide resources; it may vary also because of differences within the school in the allocation of resources and teachers to particular groups of pupils (Dyer, 1972; Elmore, 1975; McDermott, 1976; McDonald, 1976; Rakow, Airasian and Madaus, 1978; Smith, 1976).

Parents are aware of within school differences and often expend a great deal of effort to place their child with a particular teacher. The reasons for preferring one teacher over another vary of course. It may be that a teacher has the reputation for warmth, or for concentrating on basic skills, or for running an open classroom, or for being an excellent disciplinarian. Teachers *are* different, but the effect of such differences on pupils is lost in a school-wide index of average teacher experience or quality.

Similarly, individual pupils in a school are differentially exposed to school resources and facilities. For example, students in certain tracks or programs never use the language

and/or science laboratory; some never take books out of the library. The actual cost of educating different pupils in school varies according to the materials, resources and facilities used in different tracks or programs. An index of the average per pupil expenditure for the school fails to take such variation into account. Few pupils are exposed to the school's average expenditure, resources, or teacher. Clearly, inferences cannot be made about the effects that differences in the allocation of independent variables to pupils within a school have on pupil achievement when data are aggregated across pupils or pupil sub-groups to the school level (Smith, 1972).

The problems associated with the level to which data are aggregated may be more critical at some levels of education than at others. While the failure to consider the effect of within school differences in the application of resources is characteristic of most school effectiveness studies, it is at the secondary school level that this approach probably produces its most serious underestimates of the relationship between resources and pupil achievement. It is in the American high school that the differentiation of pupils into tracks or courses of study is most pronounced. The instructional goals differ sharply among various courses and tracks. Within school groupings also exist at the elementary school level, as noted above, but generally, elementary school instructional goals are not as sharply differentiated as at the high school; for example, all elementary school teachers try to instill basic reading skills in all pupils regardless of their assigned reading group.

As Table 1 indicates, the Coleman Report investigated a generalized, school-level effect upon pupils' achievement. By deciding to aggregate school resource and facility variables to the school level, the authors were unable to match particular pupils with their teachers in order to examine the relationship between pupil achievement and resource allocation within schools. This practice is not unusual, and independent variables associated with teacher characteristics and school resources have been aggregated to the school level in the overwhelming majority of differential school effectiveness studies (Coleman, *et al,* 1966; Comber and Keeves, 1973;

Husen, 1967; Jencks, *et al,* 1972; Levin, 1970; Mayeske, *et al,* 1972; Smith, 1972; Thorndike, 1973). With the focus of the researcher upon effectiveness at the global, school-wide level, data have rarely been collected to enable matching pupils with their particular classroom teachers. The issue, of course, is how one conceptualizes the effect of school resources on pupil achievement. If the school-wide average of various resources and facilities is conceived to be the most powerful influence on pupil achievement it is appropriate to aggregate data to the school level. If, on the other hand, within school influences such as the characteristics of particular teachers, texts, or resources are viewed as most influential, then a different level of data aggregation is called for (Wiley and Bock, 1967).

Table 1 shows that Mayeske, *et al* (1972) aggregated all variables to the school level. Using a dependent variable formed on the basis of a principal component analysis of all of the tests employed in the original survey and without stratifying the data by region or racial ethnic group, Mayeske, *et al* (1972) found that about 35 percent of the achievement variance lay between schools.

Other studies have examined the effect of studying within school aggregates of school effectiveness. Frederiksen (1975) reanalyzed portions of the Coleman, *et al* (1966) data by stratifying pupils within schools into eight subgroups according to race and home background variables. After conducting separate analyses for these subgroups, Frederiksen reported that different schools evidenced different achievement patterns for different types of pupils. The variance in school achievement which was associated with home background variables was only 6% for poor, black pupils, while for middle class white pupils, 81% of the between school variance was accounted for by home background variables. Frederiksen concluded that the school characteristics that are effective for one subgroup of pupils may not be the same as those that are effective for other subgroups of pupils and, "in our subsequent analyses, we must not assume that regression surfaces. . . are the same (i.e., homogeneous) for all subgroups of pupils enrolled in a school" (Frederiksen, 1975, p. 29).

A more direct examination of the differences among classes within a school involved a reanalysis of the American data from International Study of Mathematics Achievement (Rakow, Airasian and Madaus, 1978; Husen, 1967). The sample consisted of 249 American schools which had two or more mathematics teachers at either grade eight, ten or twelve. Within school variance was partitioned into a "true" error component of variance and a component attributable to the differences among classes within a school. The IEA mathematics tests appropriate to the grade levels studied were analyzed. Two main findings appeared. First, the proportion of between school variance observed on the achievement tests was considerably higher than that found in the Coleman Report, ranging from 25% at grade 8 to 41% at grade 12. Undoubtedly, the content validity of the mathematics test (recall that the Coleman Report used a verbal ability measure as the dependent variable), the fact that schools were unstratified in analysis and the increased self-selection of mathematics students at the higher grade levels largely account for these high percentages of between school variance.

Secondly, and more to the point regarding the level of data aggregation, the analyses showed that at each grade level, between 35 and 40 percent of the total within school variance was associated with differences between classes within schools. Thus, up to two-fifths of what would be labeled error or individual pupil variance if the achievement data had been aggregated to the school level was, in fact, related to differences in math achievement between classes within schools. It is important to note also that when interest in mathematics or socio-economic status was used in place of achievement as the dependent variable, only about 10% of the within school variance could be attributed to differences between classes within schools. Thus, the data aggregation problem appeared to be most critical when the dependent variable was a measure of pupil achievement.

A replication of the Coleman Report conducted in Ireland (Madaus, Kellaghan and Rakow, 1975) used both norm-referenced standardized tests and Irish external examinations (Madaus and Macnamara, 1970) to examine

differential school effectiveness. A national sample of 50 boys' schools, enrolling pupils who were preparing for the Irish Intermediate and Leaving Certificate Examinations were studied. Independent variable data on a number of school, pupil and family variables were obtained. Two types of dependent variable information were gathered: (1) pupils' performance on standardized ability and achievement tests and (2) pupils' performance on 14 intermediate and 12 leaving certificate external examinations. The external examination data were gathered one and two years respectively after the independent variable data had been gathered.

A direct comparison at the leaving certificate level (i.e., senior year in high school) was made of the percentages of between group variance obtained when data were aggregated to the school and class within school levels respectively. These comparisons were made using pupils' performance on standardized, norm-referenced tests as the dependent variable. On the total IQ score obtained from the *Primary Mental Abilities Test* (Thurstone and Thurstone, 1947), for example, 20.6 percent between group variance in pupil performance was obtained when data were aggregated to the school level as opposed to 33.3 percent when data were aggregated to the class within school level. Over a variety of standardized tests, data aggregated to class within school level revealed 12 to 17 percent more between group variance than when data were aggregated to the school level (Madaus, Kellaghan and Rakow, 1975). By senior year, high selectivity in the Irish system had reduced the variation in the pupil achievement and ability levels existing between schools and classes. This fact suggests that the differences in between group variance observed when class or school was used as the level of data aggregation probably represent lower bounds for educational systems where there is less selectivity than in Ireland.

The preceding discussion illustrates that the level to which data are aggregated can have a sizable effect on conclusions about differential school effectiveness. Different aggregation strategies produce different observed proportions of between group variance. It is interesting to speculate as to how much variance Coleman, *et al* (1966) or Mayeske, *et al*

(1972) might have found between classes at the elementary level or between tracks at the secondary level if they had been able to dis-aggregate their data to these levels. More importantly, from the perspective of integrating the findings of various studies into an interpretable form, is the fact that different levels of data aggregation imply different conceptualizations about the nature of school effectiveness. As Cronbach (1976) has pointed out, the same data studied at different levels of aggregation likely are measuring entirely different constructs. Two studies which employ dissimilar data aggregation strategies contain different conceptualizations about the process of schooling and its effects on pupils. Recognition of the different conceptualizations implied by this methodological parameter can do a great deal to help reviewers reconcile disparate, often confusing and contradictory findings appearing under the general rubric of school effectiveness research.

Selection of a Dependent Variable

Like other methodological decisions, the selection of the dependent variable in a study of school effectiveness contributes to the conceptualization of research studies. In effect, the dependent measure defines the schooling outcome about which inferences of "effectiveness" are made. Prior to 1975, the dependent variable used in all major school effectiveness studies was a standardized, norm-referenced ability and/or achievement test (Klein, 1976). There were, of course, logical reasons for this choice: such tests had a long and venerable history, they were nationally normed and standardized, and were readily available at a relatively low cost. While such tests admittedly overlooked important objectives of the school — particularly affective ones — they were regarded as valid measures of reading, computation and other verbal skills; recall that a principal criticism of schools had been that many pupils failed to learn these basic skills. Further, the relatively low performance on such tests of inner city children had been used to justify the funding of many compensatory programs, so that the selection of

standardized tests to measure school effectiveness was not surprising (Madaus, Airasian and Kellaghan, in press).

The Coleman Report, as Table 1 shows, used a general verbal ability test and found proportions of variance on the order of ten percent to exist between schools. These proportions were, however, larger than those associated with the standardized achievement tests used in the Coleman Report. Mayeske, *et al* (1972) used a composite dependent variable more heavily loaded on achievement than Coleman, *et al's* (1966) verbal ability measure, and found that about 35 percent of the variance lay between schools. However, as Table 1 indicates, Mayeske, *et al* (1972) also altered a number of other methodological parameters in their study, including the level of data aggregation.

Rakow, Airasian and Madaus (1978) found between school variance ranging from 25 to 40 percent, depending upon the grade studied, when the IEA mathematics achievement tests were used as dependent variables. Madaus, Kellaghan and Rakow's (1975) study found approximately the same proportion of variance lying between classes for their standardized ability and achievement tests. Given the high correlation between standardized measures of verbal ability and achievement it is not surprising that the amounts of observed between school variance are relatively similar for both types of measures.

The issue of selecting a dependent variable for studies of school effectiveness also may be considered at a more theoretical level. In seeking to examine school effectiveness, researchers strive to relate variation in independent variables characterizing schools to variation in a dependent measure of pupil performance. The search for such relationships is predicated on the assumption that the independent and dependent measures are construct valid. Thus, for a given dependent variable there should be a logical connection between what is taught in schools and the inferences about school achievement that are made on the basis of the measure (Klein, 1976). The construct validity of variables which are surrogates for non-instructionally related achievement or which, to a large extent, confound school with extra-school factors is open to serious question. In order that the

tests used as dependent measures in studies of school effectiveness be construct valid, that is, provide information about *school* effectiveness, they should be closely related to instruction. Concomitantly, they should be unrelated, as far as possible, to home background, social class or general verbal ability. It is, of course, impossible to completely disentangle school performance from home factors, but inferences about school effectiveness should be based on measures of effectiveness that are related as closely as possible to skills and content directly taught in schools.

It is true that standardized achievement tests — and to a lesser degree ability tests — possess a certain face validity vis a vis school outcomes. That is one reason why they have been so widely used. However, the assumption of a congruence between standardized tests and what is taught in schools is not sufficient to establish the validity of such tests for measuring school effectiveness. It is necessary to look more closely at this assumed congruence. When we do, we find that there are many reasons for questioning this assumption (Airasian and Madaus, 1976; Madaus, 1977). In the following discussion, we shall concentrate upon the case of standardized achievement tests, since these measures appear to possess the greatest face validity vis a vis school instruction. Many of the general points made in the context of achievement tests are, however, also true of standardized verbal ability tests of the type used in the Coleman Report.

For a standardized achievement test to be commercially viable and also to be fair across diverse school curricula, it must be built to tap general skills and abilities. In order to maintain fairness across diverse American school curricula, standardized achievement tests focus upon the skills and content areas that are common to the most widely used curricula. These common skill content areas are related to, but not necessarily isomorphic with, the specific instructional emphases in particular schools or classes (Shoemaker, 1975); general achievement tests tend to tap correlates of learning, rather than short term, instructionally specific learning (Madaus, Airasian and Kellaghan, in press; Stake, 1972; Stodolsky, 1972). Such achievement measures load highly on 'g' and verbal ability, variables which in turn are influenced

greatly by home background and social class factors. As a consequence, variation in standardized achievement measures tends to be more highly related to pupil's home and social class characteristics than to differences in instructional emphases or resources among schools.

Schools teach specific skills or content topics, hoping of course that these specific outcomes transfer to more generalized abilities. However, actual instructional variation occurs at the level of the component skills which comprise more general traits like reading comprehension and mathematics computation. Teachers do not teach reading comprehension *per se.* Rather, they teach a series of specific skills (sounding out words, finding topic sentences, etc.) which when integrated by the student, make up the general trait we call reading comprehension. Further, there is often great variation from school to school and from class to class within a school in the emphasis accorded particular skill areas or content topics, even when these are defined and taught the same way. These factors mitigate the usefulness of general standardized achievement tests as the dependent measure in studies of school effectiveness.

Another factor that detracts from the validity of standardized achievement tests is that such tests typically are constrained to objective, multiple choice, machine scorable response formats. In many cases, the multiple choice item format is an indirect measure of school achievement insofar as it represents a disjuncture between the way a skill was taught and the way it was measured. The method used to measure an instructional task can affect inferences about the performance level of pupils (Anderson, Scott and Hutlock, 1976; Madaus and Rippey, 1966; Stodolsky, 1972). A long standing and often ignored principle of education is that the more direct the measure of instruction, the more valid that measure. If an indirect measure has to be used, it should be validated in terms of its relationship to some direct measure of the skill or objective in question (Smith and Tyler, 1942).

When pupils are called upon to answer questions they have not been directly exposed to in school, or questions which are phrased in different ways than they are used to seeing, or questions requiring a different response mode from

that which they are familiar with, then they must rely upon a more general set of abilities in responding. It is in this circumstance that home and social class background characteristics tend to come into play.[1] If pupils cannot rely upon school specific instruction to provide a correct answer, they must rely upon non-school experiences. As soon as pupils do this, measured achievement tends to become highly loaded on non-school, home background and social class factors. Pupils with a wide range of non-school experiences and knowledge to draw upon are more likely to respond correctly than pupils from a less educationally rich background. Within school variance increases in line with the individual difference variance in social class or home background characteristics of pupils in the school.

If this argument is correct then specific differential effectiveness associated with schools or programs is most likely to be manifested on the particular skill areas contained in a standardized test, since it is at this level that variance in instructional emphasis lies. Pupils in different schools may perform differently on different skill areas comprising the total test in line with variance in instructional factors in these skill areas. Standardized tests aggregate performance on a variety of these skills to produce a total test score index, which may mask or cancel differences between pupils, classes or schools on the narrower, item specific component skills. In effect, seeking to define school or program related differences at the total test score level may result in comparisons that are insensitive to more specific school effects (Airasian, 1977; Airasian and Madaus, 1976).

The preceding argument suggests a different conceptualization of school effectiveness than that evidenced in the Coleman Report. By choosing a standardized norm-referenced verbal ability test as the criterion measure of

---

[1]Here we may consider the case of a general ability test such as that used by Coleman, *et al* (1966). Such a test, at least in theory, is designed not to load heavily upon specific in-school experiences. In this light, it is not surprising that Coleman, *et al's* (1966) use of a standardized ability test as their dependent variable produced high within school individual difference variation. Nor is it surprising that the vast majority of between school variance was explained by home, not school, variables.

school effectiveness, Coleman, *et al* (1966) made two important implicit assumptions about the cognitive outcomes of schooling. First, they assumed that the trait measured by their dependent variable (i.e., verbal ability) was an important objective of the instructional process in schools, both at the global level and in particular classes and school tracks. Second, they believed that their measure of verbal activity was sufficiently independent of home background factors to be a sensitive measure of differential school effectiveness. In essence, they assumed that a measure designed to maximize differences between individuals was suitable as well for identifying differences between groups of individuals (Carver, 1974, 1975). After the fact, the authors of the Report explicitly denied the validity of these assumptions by admitting that the skills comprising verbal ability were only incidentally taught in schools.

Recent studies in England and Ireland (Brimer, et al, 1977; Madaus, Kellaghan and Rakow, 1975) as well as in the United States (Airasian and Madaus, 1976; Brickell, 1976; Stodolsky, 1972) have conceptualized school effectiveness rather differently. In these studies, the influence of instruction on pupils was regarded as less global and more skill and content specific, thereby leading the researchers to seek dependent measures that were more closely related to the instructional process. Shoemaker (1975) and proponents of domain referenced testing (e.g., Hively, Patterson and Page, 1968) have made a similar argument. Two approaches have been followed. The first involves analyzing items within existing standardized norm-referenced tests to find item subsets which significantly differentiate between classes, schools or programs. The second approach has identified contexts where there was strong pressure on teachers and pupils to make their instruction conform to external examinations. Both approaches have resulted in inferences about differential school effectiveness which are quite different from those of the Coleman Report.

The first approach to identifying outcome measures which were more sensitive to instruction was suggested by Lewy (1973). He used the intraclass correlation coefficient (Haggard, 1958), an index of between group variation, to

identify items from a pool of arithmetic items which significantly differentiated between schools. Traditional item analysis techniques allow the user to select items which maximally differentiate between individuals regardless of their capacity to differentiate between different instructional programs, schools or classes within schools. Lewy grouped items with statistically significant intraclass correlations into a test and used the scores on this new test as the criterion variable to estimate between school achievement differences. He found substantial differences between schools which were not revealed when tests constructed through traditional item selection techniques were used as the criterion measure.

In an extension of Lewy's work (Airasian and Madaus, 1976), the items comprising a number of different standardized norm and criterion referenced tests were analyzed through Lewy's intraclass correlation method. When the items comprising a variety of norm, objective and criterion referenced tests were analyzed separately in the larger study, it was found that items which differentiated school performance did not all differentiate in the same direction. That is, some schools did exceptionally well on some items but poorly on others, while other schools evidenced the opposite pattern of performance. When the total test score was used to estimate between school variance, the differential performance of schools on the item subsets cancelled one another and resulted in negligible amounts of between school variance (Airasian and Madaus, 1976). When the many unidirectional item subsets defined from the total set of test items were used separately as criterion measures to estimate between school variance, between 5 and 10 percent of the variance lay between school on *each* relatively short (5-10 items) subtest. When these subtests were analyzed multivariately through a discriminant function analysis, between 20 and 80 percent of the variance was found to lay between schools. These findings lend support to the argument that there are differences between schools in instructional outcomes that are masked when the more global total score from a standardized test is used as the criterion variable. It is also worthy to note that subtests of items which differentiated between schools were found regardless of whether the

test examined was norm-referenced, objective-referenced, or criterion referenced. At the total test score level, norm, objective and criterion referenced tests were equally insensitive to detecting differences in achievement that lay between schools (Airasian and Madaus, 1976). These findings have particular import for the current discussion as to whether norm or criterion referenced tests are more suitable for school and program evaluation.

These results show that a change in the conceptualization of school effectiveness, as manifested by a change in the dependent variable used to define school achievement differences, can affect substantially the proportion of between school variance observed in school effectiveness studies. Different findings, of course, can substantially alter inferences made about the differential effectiveness of schools.

The second approach to identifying more sensitive measures of school related achievement involved the search for contexts in which there is a great deal of overlap between the specific topics of instruction and the specific topics measured by some test. While there are isolated examples in which American researchers have built or defined instructionally sensitive dependent variables (Brickell, 1976; Kamii, 1971), the external examination systems in Ireland (Madaus and Macnamara, 1971; Madaus, Kellaghan and Rakow, 1975) and England (Brimer, *et al,* 1977) have provided a better example of the sought for isomorphism between instruction emphasis and achievement tests.

In Ireland and England, an external agency (the Department of Education in Ireland or University-based examination boards in England) sets a syllabus of instruction for each subject taught at the post-primary school level. The same external agency also sets examinations in these subjects which are used to certify successful completion of a particular level of post primary education. Since these examinations have powerful consequences for pupils in terms of certification, advancement to third level education, and access to white collar occupations, there is great pressure on schools and teachers to focus instruction upon the syllabus of instruction which forms the basis for the external exams. The import of these exams is a powerful spur in gearing

instruction toward the content and form of the examinations. Overall, a complex of pressures assures that instruction will closely mirror the skills, knowledge and processes tested by the external examination.

The replication of the Coleman study in Ireland (Madaus, Kellaghan and Rakow, 1975) outlined in Table 1, and a similar replication in England (Brimer, *et al,* 1977) utilized external examination results as dependent variables. As noted previously, the Irish replication also used standardized achievement and ability tests as alternative dependent variables. Using the class within school as the level of data aggregation, three hypotheses were tested. The first stated that more between class variance would be associated with course-specific achievement measures (public examinations) than with the standardized tests. This hypothesis was not confirmed; both types of tests evidenced approximately the same percents of between class variance, about 30 to 35 percent, depending on the subject area examined. The second hypothesis stated that greater percents of between class variance would be associated with external examinations in subjects which have a large school input (e.g., mathematics and science) than in subjects which are less school dependent such as English or social studies. A subsidiary hypothesis was that the more specialized or advanced the subject tested by the public examinations, the greater the between class variance. The analyses confirmed both these hypotheses. As other research has shown (Airasian and Madaus, 1976; Coleman, 1975), the magnitude of between group differences is not solely a function of the type of test used, but also of the subject area and level examined; in general, the more school dependent a subject, the larger the observed percentages of between class variance. Essentially, schools seem to have their biggest differential impact on subjects which are heavily dependent upon school factors and which are less dependent on home background factors such as general verbal ability.

The third hypothesis in the Irish replication is most critical to the discussion of dependent variable selection and the effect of different types of dependent variables on conclusions about school effectiveness. The hypothesis stated

that independent variables associated with the class or school would explain higher percentages of between group variance on the course-specific, external examinations than on the standardized tests. A series of commonality analyses confirmed the hypothesis. Variables grouped into a CLASS block (emphasis on homework, class climate, etc.) and an INDIVIDUAL/CLASS block (academic stream, type of class) uniquely explained larger proportions of between class variance on the public examinations than on the standardized tests, despite the fact that the total amount of between class variance was approximately equal for the two types of dependent variables. For more than half of the 26 public examinations analyzed, over 70 percent of the between class variance was explained by the CLASS or INDIVIDUAL/CLASS blocks, with the former block being substantially more important than the latter. Unlike the Coleman Report and other studies which used standardized measures of school outcomes, very small proportions of variance were explained by home background variables. In virtually all analyses which used the public examinations as the criterion measure, home background independent variables accounted for only about two percent of the between class variance.

A comparison of the dependent variables examined in these studies with the dependent variable utilized in the Coleman Report, clearly indicates the presence of different conceptualizations of the schooling process and its observable effects on pupil achievement. The standardized verbal ability measure selected for use in the Coleman Report carries with it the implication that school instruction exerts a generalized, trait level effect on pupil learning. The use of more instructionally specific measures implies a view of school learning that is specific and focused upon demonstrable classroom instruction on specific content topics. Given the different conclusions these studies reached about the relationship of school factors to pupil achievement differences, it is clear that the dependent variable is a crucial factor in determining the validity of inferences about differential school effectiveness. Different conceptualizations of what are

reasonably expected school outcomes lead to different conclusions about the relationship between specific school factors and pupil achievement.

Selection of Independent Variables

A final methodological parameter which serves to define the conceptualization of school effectiveness is the composition of the independent variables employed. In school effectiveness studies, independent variables typically include various measures of school resources and facilities, teacher characteristics, and pupils' home background. We have already discussed how the decision to aggregate and analyze such variables in different ways can influence inferences made about school effectiveness. Similarly, the presumed and actual relationships between the independent variables and the dependent variable influence the inferences made about the differential impact of school resources on achievement.

For the most part, as Table 1 shows, the Coleman Report relied upon independent variables which assessed status factors associated with pupils, teachers, curricula and schools. Age of textbooks, years of teaching experience, presence of a science laboratory, and father's income are illustrative of the types of independent variables surveyed. Variables such as average number of science courses taught, proportion of students who come from homes owning an encyclopedia, or teacher's average salary are status independent variables and do not necessarily reflect differences in class climate or instructional process, and thus may be unable to characterize the educative process as it actually transpires in schools and classes.

Most of the Coleman Report's independent variables may be regarded as proxies for variables which are more closely related to student performance, but which are difficult to measure (Kellaghan, 1976). For example, the obtained measure of teacher's verbal ability is presumably a proxy for such variables as the teacher's ability to communicate effectively, the teacher's adequacy as a verbal model, etc. Similarly, the independent variable "teacher

experience" likely was intended to be a proxy for a variety of teacher management and instructional behaviors. Status variables are often poor proxies for the more proximate instructional or process variables. A variable such as number of books in the school library is presumably less important to pupil learning than the selection of books and their accessibility to pupils (Rivlin, 1971; Michelson, 1970). Similarly, a measure of teacher experience or verbal ability is a weak proxy for a teacher's sympathy, confidence in her students, or ability to motivate learning (Rivlin, 1971). A teacher's years of service do not indicate the discipline or press for academic excellence which characterize the teacher's classes.[1]

When school resources and facilities are defined in terms of such status variables and the dependent variable is a measure of verbal ability or a standardized achievement test, very little of the variance between schools can be explained by school factors (Coleman, *et al,* 1966; Averch, *et al,* 1972; Benson, *et al,* 1965; Goodman, 1959; Jencks, *et al,* 1972; Mollenkopf and Melville, 1965; Purves and Levine, 1965). The conceptualization that static measures of school facilities and teacher characteristics are related to school achievement differences rarely has been validated.

Several recent studies of school effectiveness have conceptualized school and teacher independent variables in a different manner. Essentially these studies, through the selection of their independent variables, suggest that the processes through which people interact with their environment are more critical for achievement than are the physical features of the environment although the physical features can set limits on the types of interactions which occur (Bloom, 1976; Brickell, 1976; Brimer, *et al,* 1977; McDonald, 1976; Madaus, Kellaghan and Rakow, 1975; Stodolsky, 1972). Certainly as far as home background characteristics are concerned, measures of environmental processes and interactions have been more powerful in

[1]Parenthetically, it should be noted that in addition to the normal measurement error expected in any variable, proxy variables contain an additional error component, since the variable one is measuring is in itself intended to be a measure of another, more basic variable.

predicting pupil performance on standardized ability and attainment tests than have measures of home and parent status (Bloom, 1964; Dave, 1963; Linnan and Airasian, 1974; Walberg and Marjoribanks, 1976; Wolf, 1964). Correlations between status indices of socio-economic status, parent's occupation or parents' education and measures of achievement tend to be .4 or less, while correlations of .7 and greater have been found between achievement and such home process variables as parents' strategies of learning reinforcement, mother's language model, and enforced sense of space and time in the home (Bloom, 1964; Fraser, 1959).

The resources and facilities offered to pupils by a teacher, a school, or a track within a school may be thought of as being comprised of two different types of variables. On the one hand are the books, expenditures, laboratories, etc., — status variables — which have occupied the attention of most school effectiveness researchers. On the other hand are a series of less tangible attitudinal or climatic variables such as discipline, press for high achievement, and directness of instruction vis a vis academic goals. These latter, process-type variables define the educational press of the school, track or classroom. They define the impetus for teachers and pupils to use the school's resources and facilities in striving for high achievement.

When the schooling process is conceptualized in this manner, it follows that school quality is a function of both types of variables, the opportunities for learning in the form of resources and facilities *and* the press to use these resources and facilities. Virtually every study which has examined the relationship between physical facilities and resources and school achievement has documented that the static, tangible characteristics of schools alone are not highly related to pupil achievement as measured by standardized norm referenced achievement or ability tests. However, when measures of the press, climate or practices which characterize a school or a class within a school are included among the independent variables surveyed, school or class specific variables explain greater proportions of between group variance on achievement measures which are more specific to the instructional process than when status

independent variables alone are used (Bloom, 1976; Brickell, 1976; Brimer *et al,* 1977; Cadzen, 1971; Kamii, 1971; McDonald, 1976; Madaus, Kellaghan and Rakow, 1975; Stodolsky, 1972).

Table 1 shows that the Madaus, Kellaghan and Rakow (1975) study revealed that the proportions of between class variance obtained were similar for standardized, norm-referenced tests and for the various external examinations investigated. However, the proportion of the between group variance which could be accounted for by class and school factors was much more for the external examinations than for the standardized tests. Thus, accounting for between group variance with particular types of independent variables cannot be understood without an adequate conceptualization of the dependent variables used to define the between group achievement differences.

Hence, while it is encouraging that various school and class presses have been related to group achievement differences, these findings must be considered within the context of at least one caveat. It must be realized that school press variables have been investigated solely within the confines of a limited type of output measure. That is, the presses found to differentiate between school performance have been found to do so only for cognitive achievement measures that were specific to the class or school's instructional emphases. Essentially, when the type of presses extant in a school or class are congruent to the types of outcomes measured on cognitive achievement tests, press variables appear as important contributors to school achievement differences. There is, in essence, a congruence in the conceptualization of the independent and dependent variables studied. If school effectiveness were conceptualized differently to include, say, affective outcomes, it is not at all clear that similar press variables would be related to observed school differences. It would be necessary to view more closely the conceptualization which related the independent and dependent variables in such studies.

When cognitive achievement measures are used as the dependent variable in studies of school effectiveness, a variety of studies suggest that a cluster of press variables generally

associated with a 'traditional' education approach, is powerful in explaining between group achievement variance (Bloom, 1976; Brimer, *et al,* 1977; McDonald, 1976; Madaus, Kellaghan and Rakow, 1975; Smith, 1976). Among these variables are (1) the extent to which the pupil actively participates in learning activities; (2) the amount of feedback and diagnosis given the pupil about his progress towards meeting well-defined academic goals; (3) the extent to which teachers and pupils conform to an agreed upon academic press and rules of discipline; (4) the "shared purposefulness" of the school or class, that is, the extent to which teachers, pupils, parents and administrators concur on the important ends of education and move towards these in a direct, purposeful manner; (5) the directness of instruction or the degree to which topics emphasized reflect the desired ends of instruction; and (6) the amount of instructional time devoted to teaching valued objectives.

In general, students who perform well on an instructionally specific achievement test are those who expect to continue their education beyond secondary school, who see themselves as hard working and scholastically able and who are more concerned with academic matters than with social popularity or participation in extra curricular activities (Madaus, Kellaghan and Rakow, 1975). Classes, as opposed to individual students, which score highly on dependent measures are those which are characterized by high educational aspirations and conformity to a well defined academic press. Large amounts of class time are spent in study and discipline is strict (Brimer, *et al,* 1977). As Madaus, Kellaghan and Rakow (1975) note, "All this suggests a climate in which there are social rewards for academic excellence, in which scholastic achievement is valued by teachers and student body" (p. 773). When the process of schooling is conceptualized to include variables of this type, as opposed to more static measures of resources and facilities, large proportions of between group achievement variance can be accounted for by school and class influences.

It has been argued that process-type variables are not as amenable to policy manipulation as are status variables which tap school resources and facilities (Cain and Watts,

1970; Coleman, 1975). While this assessment may be true given the present state of knowledge about the schooling process — it clearly is an easier matter to increase per pupil expenditure in a school than it is to heighten academic press — it is a rather short-sighted position to adopt. If one aim of evaluative research is to improve social processes and institutions, then a more accurate understanding of these processes is critical. In light of this goal, it hardly seems justifiable to limit the variables one examines to those most amenable to manipulation. Given our present knowledge base about schooling and the factors related to differential school effectiveness, it seems most essential that we focus our attention upon the conceptualizations adopted in different studies and endeavor to critique the validity of the results obtained under these varied conceptualizations.

Conclusion

Throughout this paper we have suggested that the methodological parameters of evaluative research studies help to identify a conceptualization of the phenomenon under investigation. We have examined the Coleman Report's conceptualization of school effectiveness and shown how changes in this conceptualization, as defined by changes in a number of methodological parameters, alter findings about differential school effectiveness and its antecedent factors. While the effects of data analysis strategy, stratification levels, level of data aggregation, and selection of independent and dependent variables are considered separately in the body of the paper, it was clear that the separate impact that each of these parameters had on conclusions about school effectiveness was often difficult to discern. Most frequently, the unique and joint effect of two or more of these parameters were at the root of altered findings. Of course, this observation is logical, given that the parameters taken together represent the researchers' conceptualization of schooling. What is most important, however, is that discussion of the methodological parameters characterizing varied research studies be put in a more general, conceptual framework, so that the model of

schooling being examined by different researchers can be identified and critiqued. Without an awareness of the model underlying an evaluative research endeavor, criticism of the study tends to be piecemeal.

Table 1 presented a brief overview of three studies of school effectiveness. Differences in the methodological parameters of these studies and the impact of these differences on conclusions about school effectiveness were identified. From this table a conceptualization of schooling and school effectiveness was defined for each study. It was interesting to note how these different conceptualizations have led to quite different conclusions about school effectiveness and its antecedent factors.

People often wonder why one consistent result of social science research is to beget more social science research. Why is it, they ask, that we get more questions than answers from expensive research endeavors. One reason is that results continually tell us that our simplistic representations of social processes are inadequate; that more is not necessarily better; that what goes into the black box is less important than what goes on inside it. Research seeks to simplify human affairs, but it most often has the opposite effect, revealing them to be more complex than originally imagined. In one sense, research has become the tail that wags the dog by continually leading us to more complex conceptualizations of social phenomena and process, rather than bringing closure to questions as initially posed.

It is often said that if you want to understand something, try to change it. We might also say, in conclusion, that if you want to change something, try to understand it. In trying to understand the processes and phenomena which are the domain of evaluative research we must do more than critique studies for their specific methodological shortcomings. We must critique such studies in a more holistic sense so that we are aware of the conceptual view embodied by the research methodology and thus have a context within which to compare the results of different studies. Evaluative research is, at heart, a test of such conceptualizations.

References

Airasian, P. W. Examining school effectiveness. Fourth invited lecture in the 1977 Spring Colloquium Series. Department of Education, University of Minnesota, Minneapolis, Minnesota, June 2, 1977.

Airasian, P. W. and Madaus, G. F. A study of the sensitivity of school and program effectiveness measures. Report submitted to the Carnegie Corporation of New York, July, 1976. Chestnut Hill, Mass.: Boston College 1976.

Airasian, P. W. and Madaus, G. F. A post-hoc technique for identifying between program differences in achievement. *Studies in Educational Evaluation,* 1978, *4,* 1, 1-8.

Anderson, L. W., Scott, C. C. and Hutlock, N. The effects of a mastery learning program of selected cognitive, affective and interpersonal variables in grades 1 through 6. Paper read at Annual Meeting of American Educational Research Association, 1976.

Armor, D. J. School and family effects on black and white achievement: a re-examination of the USOE data. In Mosteller, F. and Moynihan, D. (Eds.), *On equality of educational opportunity.* New York: Vintage Books, 1972, 169-229.

Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J. and Pincus, J. *How effective is schooling? A critical review and synthesis of research findings.* Santa Monica, California: Rand Corporation, 1972.

Beaton, A. E. Commonality. Unpublished manuscript, 1973.

Beaton, A. E. Multivariate commonality analysis. In Mayeske, G. W., *et al.,* Technical Supplement to *A Study*

of the achievement of our nation's students. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1974.

Benson, C., *et al. State and local fiscal relationships in public education in California.* Report of the Senate Fact Finding Committee on Revenue and Taxation, Senate of the State of California, Sacramento, March 1965.

Bloom, B. S. *Stability and change in human characteristics.* New York: Wiley, 1964.

Bloom, B. S. *Human characteristics and school learning.* New York: McGraw-Hill, 1976.

Bowles, S. and Levin, H. M. The determinants of scholastic achievement: An appraisal of some recent evidence. *Journal of Human Resources,* 1968, *3,* 3-24.

Brickell, H. M. "Needed: instruments as good as our eyes." *Evaluation Center Occasional Paper Series #7.* College of Education, Western Michigan University. Kalamazoo, Michigan, 1976.

Brimer, A., *et al.* A study of the sensitivity of measures of school effectiveness in England. Report submitted to the Carnegie Corporation of New York, 1977.

Burkhead, J., Fox, T. G., and Holland, J. W. *Input and output in large city high schools.* Syracuse, N.Y.: Syracuse University Press, 1967.

Burstein, L. The choice of unit of analysis in the investigation of school effects: IEA in New Zealand. *New Zealand Journal of Educational Studies,* 1976, *11,* 11-24.

Cain, G. G. and Watts, H. W. Problems in making policy inferences from the Coleman Report. *American Sociological Review,* 1970, *35,* 228-242.

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. *American Psychologist,* 1974, *29,* 512-518.

Carver, R. P. The Coleman Report: Using inappropriately designed achievement tests. *American Educational Research Journal,* 1975, *12,* 77-86.

Cazden, C. B. Evaluation of learning in preschool education. In B. Bloom, J. T. Hastings and G. F. Madaus. *Handbook on formative and summative evaluation of student learning.* N.Y.: McGraw-Hill, 1971, 345-398.

Cicirelli, V. G., Cooper, W. H. and Granger. R. L. *The impact of Head Start: An evaluation of the effect of Head Start on children's cognitive and affective development.* Study by Westinghouse Learning Corporation and Ohio University. Washington, D. C.: Office of Economic Opportunity, 1969.

Cohen, J. and Cohen, P. *Applied multiple regression/correlation analysis for behavioral sciences.* New York: Wiley, 1975.

Coleman, J. S. Methods and results in the IEA studies of effects of school on learning. *Review of Education Research,* 1975, *45,* 335-386.

Coleman, J. S., Campbell, E. O., Hobson, C. J., McPartland, J., Mood, A. M., Weinfried, F. D., and York, R. L. *Equality of educational opportunity.* Washington, D. C.: U.S. Government Printing Office, 1966 (2 vols.).

Comber, L. C. and Keeves, J. P. *Science education in nineteen countries: An empirical study.* N.Y.: Wiley, 1973.

Cronbach, L. J. Research on classrooms and schools: Formulation of questions, designs and analyses.

Occasional Paper of the Stanford Evaluation Consortium, Stanford University. Stanford, California, 1976.

Dave, R. H. The identification and measurement of environmental process variables that are related to educational achievement. Doctoral dissertation, University of Chicago, 1963.

Dyer, H. S. The measurement of educational opportunity. In F. Mosteller and D. Moynihan (Eds.) *On equality of educational opportunity.* N. Y.: Random House, Inc., 1972.

Elmore R. F. Design of the Follow Through experiment. In A. Rivlin and M. Timpane, (Eds.) Planned variation in education. Washington, D.C.: The Brookings Institution, 1975, 23-46.

Fraser, E. *Home environment and the school.* London: University of London Press, 1959.

Frederiksen, J. R. School effectiveness and equality of educational opportunity. Report submitted to the Carnegie Corporation of New York, 1975. Cambridge, Mass.: Harvard University.

Glass, G. F., *et al.* Education of the disadvantaged: An evaluation report on Title I Elementary and Secondary Education Act of 1965, Fiscal Year 1969. Boulder, Colo.: University of Colorado, 1970 (mimeo).

Goodman, S. M. *The assessment of school quality.* Albany, N.Y.: New York State Education Department, March, 1969.

Haggard, E. *Intraclass correlation and the analysis of variance.* N.Y.: Dryden Press, 1968.

Hanushek, E. A. and Kain, J. F. On the value of *Equality of educational opportunity* as a guide to public

policy. In F. Mosteller and D. P. Moynihan (Eds.) *On equality of educational opportunity.* Papers deriving from the Harvard University Faculty Seminar on the Coleman Report, N.Y.: Vintage Books, 1972.

Hively, W., Patterson, H. L. and Page, S. A universe defined system of arithmetic achievement tests. *Journal of Educational Measurement,* 1968, *5,* 275-290.

Husen, T. *International study of achievement in mathematics. N.Y.: Wiley, 1967.*

Jencks, C. S. The quality of the data collected by The equality of educational opportunity survey. In F. Mosteller and D. Moynihan (Eds.). *On equality of educational opportunity.* Papers deriving from the Harvard University Faculty Seminar on the Coleman Report. N.Y.: Vintage Books, 1972.

Jencks, C. S., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. *Inequality: A reassessment of the effect of family and schooling in America.* N.Y.: Basic Books, 1972.

Kamii, C. K. Evaluation of learning in preschool education: socio-emotional, perceptual-motors, and cognitive development. In B. S. Bloom, J. T. Hastings, and G. F. Madaus (Eds.). *Handbook on formative and summative evaluation of student learning.* N.Y.: McGraw-Hill, 1971.

Kellaghan, T. Measuring school effectiveness. Dublin: Educational Research Centre, St. Patrick's College, 1976.

Kiesling, H. J. *The relationship of school inputs to public school performance in New York State.* Santa Monica, Calif.: The Rand Corporation, 1969, Report # P-4211.

Klein, S. P. Cost-quality research limitations: The problem of poor indices. In J. E. McDermott (Ed.) *Indeter-*

minacy in education. Berkeley, Calif.: McCutchan Publishing Co., 1976, 33-49.

Levin, H. M. A new model of school effectiveness. In *How do teachers make a difference.* Washington, D.C. U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Educational Personnel Development (OE-58042), 1970, 55-75.

Lewy, A. Discrimination among individuals v. discrimination among groups. *Journal of Educational Measurement,* 1973, *10,* 19-24.

Linnan, R. and Airasian, P. W. Ethnic comparisons of environmental predictors of three cognitive abilities. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, 1974.

Madaus, G. F. The use of standardized tests in measuring output. Paper presented at the University of Connecticut, Department of Education Colloquium, March 7, 1977 (mimeo).

Madaus, G. F., Airasian, P. W. and Kellaghan, T. *School effectiveness: A reassessment of the evidence.* N.Y.: McGraw-Hill, in press.

Madaus, G. F. Kellaghan, T. and Rakow, E. A study of the sensitivity of measures of school effectiveness. Report submitted to the Carnegie Corporation of New York. Educational Research Centre, St. Patrick's College, Dublin, and Boston College, 1975.

Madaus, G. F. and Macnamara, J. *Public examinations: A study of the Irish Leaving Certificate.* Dublin: Educational Research Centre, St. Patrick's College, 1970.

Madaus, G.F. and Rippey, R. Zeroing in on the *STEP Writing Test:* What does it tell a teacher? *Journal of Educational Measurement,* 1966, *3,* 1.

Mayeske, G., *et al. A study of our nation's schools.* Washington, D.C.: U.S. Office of Education, 1972.

Mayeske, G. W., Okada, T., Cohen, W. M., Beaton, A. E., and Wisler, C. F. *A study of the achievement of our nation's schools.* DHEW Publication No. (OE) 72-31. Washington, D.C.: Department of Health, Education and Welfare, 1973.

McDermott, J. E. (Ed.) *Indeterminacy in education.* Berkeley, Califl: McCutchan 1976.

McDonald, F. J. *Teachers do make a difference.* Princeton, N.J.: Educational Testing Service, 1976.

Michelson, S. The association of teachers resourcefulness with children's characteristics. In *How do teachers make a difference.* Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Educational Personnel Development (OE-58-42), 1970, 120-168.

Mollenkopf, W. G. and Melville, S. D. A study of secondary school characteristics as related to test scores. Research Bulletin 56-6. Princeton, N.J.: Educational Testing Service, 1956 (mimeo).

Mosteller, F. and Moynihan, D. P. (Eds.) *On equality of educational opportunity. Papers deriving from the Harvard University Faculty Seminar on the Coleman Report.* N.Y.: Vintage Books, 1972.

Purves, A. C. and Levine, D. F. (Eds.) *Educational policy and international assessment.* Berkeley, California: McCutchan Publishing Corp., 1975.

Rakow, E., Airasian, P. and Madaus, G. Assessing school and program effectiveness: estimating hidden classroom level effects. *Journal of Educational Measurement,* 1978, *15,* 1, 15-21.

Rivlin, A. M. *Systematic thinking for social action.* Washington, D.C.: Brookings Institution, 1971.

Shea, B. Schooling and its antecedents. *Review of Educational Research,* 1976, *46,* 4, 463-526.

Shoemaker, D. M. Toward a framework for achievement testing. *Review of Educational Research,* Winter, 1975, *45,* 1, 127-147.

Smith, M. Equality of educational opportunity: The basic findings reconsidered. In F. Mosteller and D. P. Moynihan, (Eds.) *On equality of educational opportunity. Papers deriving from the Harvard University Faculty Seminar on the Coleman Report.* N.Y.: Vintage Books, 1972.

Smith, M. Discussion. Evaluation of educational programs. In C. Abt (Ed.) *The evaluation of social programs.* Beverly Hills, Calif.: Sage Publications, 1976, 430-434.

Smith, E. R. and Tyler, R. W. *Appraising and recording student progress.* N.Y.: Harper, 1942.

Stake, R. E. Measuring what learners learn (with a special look at performance contracting). Urbana, Illinois: Center for Instructional Research and Curriculum Evaluation. University of Illinois, 1972 (mimeo).

Stodolsky, S. S. Defining treatment and outcome in early childhood education. In H. T. Walberg and A. T. Kopan (Eds.) *Rethinking urban education.* San Francisco, Calif.: Jossey Bass, 1972.

Thorndike, R. L. *Reading comprehension education in fifteen countries: An empirical study.* N.Y.: Wiley, 1973.

Thurstone, L. L. and Thurstone, T. G. *SRA Primary Mental Abilities Ages 11-17, Form A4.* Chicago: Science Research Associates, 1947.

Walberg, H. and Marjoribanks, K. Family environment and cognitive development: Two analytic models. *Review of Educational Research,* 1976, *46,* 4, 527-551.

Wiley, D. C. Another hour, another day. Quantity of schooling, a potent path for policy. Mehr Licht, Studies of Educative Processes, Report No. 3, July 1973.

Wiley, D. and Bock, R. Quasi-experimentation in educational settings: comment. *The School Review,* 1967, *75,* 1, 353-366.

Wisler, C. E. Partitioning the explained variance in a regression analysis. In G. W. Mayeske *et al, Technical Supplement to 'A study of the achievement of our nation's students'.* Washington, D.C.: U.S. Department of Health, Education and Welfare, 1974.

Wolf, R. The identification and measurement of environmental process variables related to intelligence. Ph.D. dissertation, University of Chicago, 1964.